International Journal of Current Engineering and Technology
E-ISSN 2277 – 4106, P-ISSN 2347 – 5161
Available at http://inpressco.com/category/ijcet

Research Article

An Overview of Medical Images Classification based on CNN

Toqa A. Sadoon* **and Mohammad H. Ali**

Department of Electronic and Communications Engineering, College of Engineering, Al-Nahrain University PO Box 64040, Baghdad, Iraq

Received 01 Sept 2020, Accepted 01 Nov 2020, Available online 03 Nov 2020, **Vol.10, No.6 (Nov/Dec 2020)**

Abstract

Classifying medical images is a fundamental problem in computer vision as well as image segmentation and detection. There are a lot of methods used to classify images such as Artificial Neural Network (ANN), Support Vector Machine (SVM) and Convolutional Neural Network (CNN). CNN has been played a vital role in classification in the last years, due to the power of deep learning among other types of machine learning. In this paper, the basic concept of CNN has been summarized. Moreover, various models are simply discussed on the CNN.

Keywords: *Medical Image, Image Pre-processing, Image Classification, Deep Learning, Convolutional Neural Network.*

1. Introduction

The evolution of the medical image in recent years has helped doctors to get more information about the human body to detect diseases. Over the years there are many types of medical images have been discovered, each with its own benefits and drawbacks. Moreover, these images can be saved in computer from medical image devices. As for the diagnosis, common types of imaging include X-ray Ultrasound, Magnetic Resonance Imaging (MRI) and Computer Tomography (CT) (LeCun, Y. *et al*, 1990). So, the medical image classification is essential to early detection of disease, Motivated by the above, there is a need to work on programs that help the doctors to detect the tumor or any disease in less time and more accurate. The researchers have been providing many research in this field such as Artificial Neural Network (ANN), Support Vector Machine (SVM) and Convolutional Neural Network (CNN). But recently the use of CNN in the classification of medical images has become very popular. ConvNet or (CNN) is a type of deep learning model that contains many layers that are used as feature extraction and image classifying, which has been discussed in the next section. In fact, CNN was introduced in the eighties and nineties (Krizhevsky, Alex, 2013). But it was not useful in the real world; therefore it was forgotten for a while. Since 2012 when it was dramatically revived (M. D. Zeiler, *et al*, 2014), until this time CNN is used in most areas of computer vision and is growing at a rapid rate. AlexNet is a CNN model designed by Krizhevky *et al* in 2012(Krizhevsky, Alex, 2013), which contains 5 convolutional layers and 3 fully connected layers.

After this model, Zeiler and Fergus, proposed ZFNet (Simonyan, K et at, 2014), which is like AlexNet except that it has a different size of filters. Then, 6 different CNN configurations A, A-LRN, B, C, D (VGG16) and E (VGG19), developed by Simonyan and Zisserman in 2014, called VGGNet (Farhana Sultana *et al*, 2018). The other models have been summarized by Farhana Sultana *et al*, in (Phil Kim, 2017). Deep learning toolbox from MATLAB providing some of these models as a pre-trained model that can be used to decrease training time, offer and give the optimal solution to the problem. Currently, the increase in the dataset used in the CNN network, the strength of deep learning frameworks and the power of the graphics processing unit (GPU) that helped design, train and validate deep neural networks, in addition to that, helped to develop many models in these years (Yann LeCun *et al*, 2015). Some of these models are well discussed in this paper. These models can be run through a high-level programming interface that relies on NVIDIA GPUs accelerated libraries. Common deep learning frameworks are: PyTorch, MXNet, TensorFlow, MATLAB, NVIDA Caffe, Chainer and PaddlePaddle.

2. Convolutional Neural Network

CNN is a deep neural network that contains many hidden layers hidden layers, these layers are used to extract the feature from the input image (Ian Goodfellow *et al*. 2017). Before ConvNet, feature extractors were independent of machine learning, it found by some other methodsand this requires a significant amount of time and cost. In contrast, when we talk about CNN, the situation is different because it includes in the training process feature extractor. Besides, the classification will perform better when the neural network of extract features contains more layers.

*Corresponding author's ORCID ID: 0000-0003-1064-6326
DOI: https://doi.org/10.14741/ijcet/v.10.6.1

In other words, (CNN) or ConvNet is a neural network that extracts input image features, the extracted feature enters into the classification model and then generates output. A simple model of CNN shown in Fig. 1.

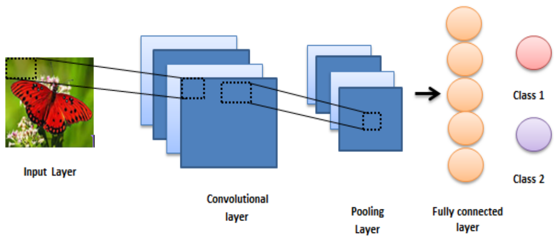

Fig.1 Simple model of CNN.

2.1 Convolutional Layer

It is the first layer in CNN that use to extract the feature from the input image. In this layer, the input image has been convolved with convolutional filters (also called kernel). A new images are the output of this layer called feature maps, which is equal to the number of a kernel that has been used in this layer (Phil Kim, 2017), each filter is responsible to extract different features from the image. Fig. 2 show an example of convolutional layer with kernel size 3x3.

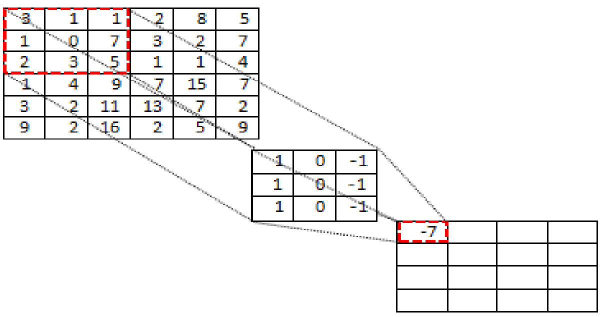

Fig.2 Convolutional Layer

2.2 Pooling Layer

The pooling layer was used to reduce the image dimensions and as a result, the complexity of CNN was reduced. The idea of the pooling layer is similar to a convolutional layer that uses the moving filter on the image and determines the result corresponding to the type of pooling layer that has been used in the model. There are many kinds of pooling layer such as, average pooling, min pooling, max pooling (most common one) and etc. Fig. 3 show an example of max pooling layer.

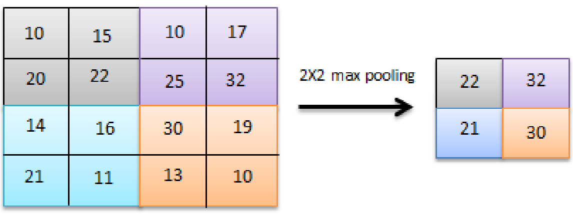

Fig.3 Example on Max-pooling Layer

2.3 Fully-Connected Layer

In this layer, the two-dimensional image turns into 1D, as well as each neuron connected to the previous neuron and performs the final result of the classification. Fig. 4 shows a fully connected layer with two categories in output.

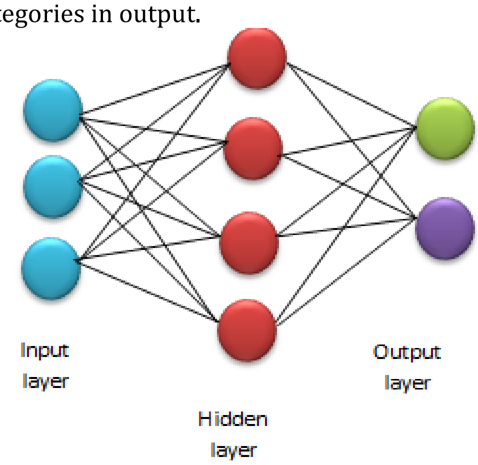

Fig.4 Fully-Connected Layer.

2.4 Dropout Layer

A common method to prevent overfitting is this layer. Dropout layer selected some nodes randomly and the number of nodes depended on a percentage value. The output of the selected nodes set to zero and as a result do not affected on the model through training process. Fig. 5 shows the nodes before and after dropout.

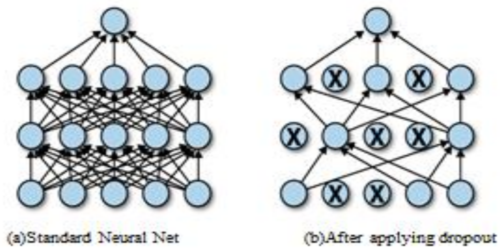

Fig.5 Dropout Layer (Reza Bosagh Zadeh *et al*. 2019).

2.5 Activation Function and Loss function

In any neural network, the activation function in a specific input node is responsible to determine the behavior of the output for this node. In deep learning including CNN, there are many activation function used such as Rectified Linear Unit (ReLU), sigmoid, softmax, Tanh and etc. In the beginning, the sigmoid activation function make the output range between zero and one and this usually used for binary classification (Phil Kim, 2017). The following equation is the mathematical representation for this function:

$$f(x) = \frac{1}{1+e^{-x}} \quad (1)$$

While, the Rectified Linear Unit activation function (Nair, V., *et al*,2010), consider a common one use in this field especially in CNN. The output of this activation is positive number and zero, this function is used to solve vanishing gradient problem through back-propagation algorithm (Phil Kim, 2017). The following equation is the mathematical representation for this function:

$$f(x) = \max(0,x) \tag{2}$$

Softmax activation function, used for determine the probabilities in the output node for each category and the biggest value of probability represented the correct class. For this reason this function used for multi-classification problems. The following equation is the mathematical representation for this function:

$$f(x_i) = \frac{e^{xi}}{\sum_{j=1}^{N} e^{xj}} \tag{3}$$

Tanh gives the output value between the range of [-1, 1]. The following equation is the mathematical representation for this function:

$$f(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \tag{4}$$

Where, x is the input and f(x) is the activation function. On the other hand, loss function used to calculate the error in the neural network, which tells us the prediction value from the actual value, then this value will be improved during the learning process again. To calculate the loss function, there are two parameters used which are the output through learning process and the correct output (label) in the case of supervised learning algorithms.

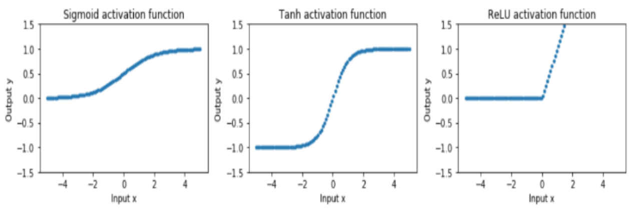

Fig.6 Activation Functions (https://h1ros.github.io/categories/pytorch/).

3. Image pre-processing and Augmentation

Pre-processing is very important step to correct, adjust and obtain non-contaminated medical image for more study and processing such as segmentation, reconstruction and classification. There are various techniques and algorithms used for this purpose and this led to decrease the complicity of the application model that uses the image and also improve the performance. The first and popular techniques are filtering, these filters are used to improve image quality, eliminate noise, preserve edges within an image and improve and soften the image. For exampe, mean filter, wiener filter and adaptive median filter (R. Ramani,*et al*, 2013). Moreover, Image preprocessing may include the bad lines restoration, detection, image registration or geometric rectification, atmospheric correction, radiometric calibration and topographic correction. Shuffle and down size image is also another techniques used for image pre-processing. As for the classification of medical images, the pre-processing step for medical image is essential, because this dealing with patient and this step is increase the performance of classification.

On the other hand, when the number of epoch is not suitable and the data set that use through learning is small, this causes the model to be overfitting. Overfitting in simple words is that, best learning performance against bad testing performance. To reduce or prevent overfitting, there are many techniques used such as data augmentation and dropout layer (has been discussed in section 2.4). Data augmentation is used to increase the data set that use in learning without losing any information and as a result reduce overfitting, also increase the performance in testing step. There are several techniques used for augmentation before the data fed into the model. For instance, rotation with different angles, scaling, flipping, mirror and shifting.

It is worth mentioning that, data augmentation solve another problem besides the above, which is that, most of hospitals do not preparing the researcher with a lot of medical image that help to complete the research and as a result, reduce overfitting and achieve good accuracy, so augmentation is a good choice for these reasons.

4. Training process and Hyper parameter of CNN

After pre-processing and data augmentation step, there are another important step in Convolutional neural network which is training processing. During training processing there are many parameter should be initialized or selected.

1) The number of convolutional, pooling, dropout and fully connected layer. Moreover, the size, stride and the number of filters that used in convolutional layer to generate feature map and we must be note that, more number of convolutional layer more feature extraction. Then, the types of pooling layer with size and stride must be selected. After that, the ratio of dropout layer and the suitable activation function must be also initialized.
2) The learning rate determines the update of weight each time. If the value of the learning rate is very high, the output fails to reaches the optimal solution. In contrast, if the value is very low, the output reaches the solution very slowly.
3) Epoch, is represented the repeated training iteration with the same dataset. On each iteration the data is goes through the learning process including updating the weight. There is no rule for determine the number of epoch, but the appropriate one places the model over the underfitting and on the bottom of overfitting which is the optimal model.
4) Optimizer methods, there are many schemes used for supervised learning in CNN to calculate the

weight update. Common schemes are, Stochastic Gradient Descent, Batch Gradient Descent and Mini Batch Gradient Descent (Phil Kim, 2017).

Firstly, Stochastic Gradient Descent (SGD) (Bottou, L., 2010), calculates the error and adjusts the weights in each training data. The benefit of using SGD is faster in case of big data set.

Secondly, Batch Gradient Descent (Ruder, S. 2016), the weight is update only once for all training data error. In contrast of SGD the batch gradient descent is more stable but slowly in the case of big data set.

Finally, Mini Batch Gradient Descent, is combining between the benefits of SGD and batch gradient descent. It selected a part of training data set and applied batch gradient descent on them. In other word, it updates the weight for selected data set (Phil Kim, 2017).

5. CNN Models for Medical Image Classification

In this section, some CNN architectures in the recently years have been simply summarize and table (1) contained the materials used in these models. The models summarize in the following sections:

1) (Hossam H.Sultan *et al*. 2019), proposed a CNN model that uses to classify MRI brain tumors. They use different data set, the first one is 3064 slices of (meningioma, glioma and pituitary tumor) from 233 patients and it contains axial, coronal and sagittal views. The second one includes 516 slices of T1-weight that include glioma grades which are Grade II, Grade III and Grade IV from 73 patients. The proposed model begins from the input layer, 3 convolutional layers, 3 ReLU layers, normalization layer, 3 max-pooling layers, 2 dropout layers, fully connected layer and softmax layer to find the final output. Before the images enter the model, there are several processing that has been made on the images to reduce the complicity as a result better performance, this processing began with reducing images size from 512x512 to 128x128and then it was shuffled. After that, they augment the first dataset to avoid overfitting and increase the number of images, some of these augmentations are, add salt noise, flipping, mirroring and rotation.
2) (Amin Kabir *et al*. 2019), use the genetic algorithm (GA) to achieve the best model performance of CNN. They chose this method to find the best CNN model, unlike other methods based on trial and error. The idea of using GA is to looking for the best parameters such as numbers of layers, numbers of filters and its size and learning rate, as a result when the best CNN has been found, GA is stopped searching when there is no improvement in the validation accuracy or when exceeds the maximum number of generation. The dataset was used in this model is the same as in [17]. All the images are rescaling into 128x128. Rotation, translating pixel, scaling and mirror used to increase the dataset. After that, the dataset was rise to 8000 normal images, 8000 glioma grades images and 1521 images for (meningioma, glioma and pituitary tumor). Moreover, there are some images were randomly selected for test purposes. The best performance for classifying glioma grades is when CNN has 5 convolutional each one followed by max-pooling layers and one fully connected layer. Furthermore, using ReLU and dropout layer. In the case of classifying the second set of images, the best network has had 6 convolutional each one followed by max-pooling layers and one fully connected layer. The authors use ELU instead of ReLU as an activation function.
3) (Ali M.HASAN *et al*. 2019), use CNN to extract the feature from 6000 MRI brain tumor images and combine these feature with extracted handcrafted features from modified gray level co-occurrence matrix (MGLCM). Then use SVM to classify the feature into normal and abnormal. For feature extraction based on deep learning, a simple CNN is used for this purpose. The model has 7 layers which are the input layer, three convolutional layers, two pooling layers and fully connected layer. Before the feature extraction, images have been pre-proposed to reduce the effect of variation and noise. Image preprocessing includes image enhancement, resizing, intensity normalization and most of the time, mid-sagittal plane detection and correction (MSP) is necessary for estimating the tumor detection. The accuracy of this method is 99.30%.
4) (S. Sasikala *et al*. 2018), presented a CNN model for classify the CT lung cancer into two main type malignant or benign. These data sets are available in Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI), which 1000 CT scans in (DICOM) format. In the pre-treatment stage, before the image is fed into the model, the median filter is used to reduce the effects of the degradation during the training and then segment the tumor. The output images are used to train CNN. The model accuracy is 96%.
5) (Gaobo Liang *et al*. 2018), presented a model that combines the pre-trained CNN with Recurrent neural networks (RNN), to classify 12,444 augmented images of blood cells in JPEG format, into four types, eosinophil, lymphocyte, monocyte and neutrophil. In the pre-processing step, the rotation matrix has been used which is crucial to the success of this method. This combination is based on use the Xception model, which is another improved model from Google Net, with a special type of RNN, which is LSTM (Long Short-Term Memory, followed by Merge layer (to incorporate features obtained from both CNN and RNN)and fully connected layer followed by Softmax as output layer.
6) (Muhammad Sajjad *et al*. 2019), use the pre-trained VGG-19, to classify the two segmentation datasets MRI brain tumor. The first datasets from Radiopaedia, which consists of 121 MR images divided into four different grades, grade I, grade II, grade III and grade IV. The second datasets is 3064 slices of (meningioma, glioma and pituitary tumor). Before the data fed into the model, firstly, the tumor is segmented using Input Cascade CNN. Secondly, apply the augmentation techniques such as, flipping, skewness, rotation, shears for geometric

transformations invariance Gaussian blur, edge detection, sharpening and emboss.. The authors measured the accuracy before and after increasing the data and the result was a demonstration that the data augmentation increases the accuracy of the system.

7) (Ashnil Kumar *et al*. 2016), presented an ensemble of different CNN method for classifying 30 modalities of medical images. These modalities contain 6776 training images and 4166 test images. The main idea of ensemble learning is to combining the output of multiple classification models and the result is one high performance classification model. The authors used two pre-trained models which is AlexNet and GoogLeNet. In ensemble learning, they used two type of classifier which is softmax and multi-class SVM. Like others models the authors also used data augmentation to reduce overfitting. Flipping and cropping for each image could increase the data from 6776 to 67760.
8) (Varun *et al*. 2019), use one of the six different architecture of VGG, which is ImageNet-VGG-f CNN. Judging from this model, the authors have been modified the VGG-f CNN by using less number of layers and six external features also have been suggested, which are fed into the CNN. Since the convolutional layers are used to extract the feature, it has been dropped in this model due to the externally extracted features. The rest of this structure is ReLU layer, max-pooling layer, fully connected layer and finally classification layer. This model is used to classify the CT images of lungs. The first datasets are three different classes which are (9 non-smokers, 10 smokers and 20 smokers with chronic obstructive pulmonary disease (COPD or emphysema)). The second data sets are ten different classes, which is also provided with the location of nodules in the patient's lungs.
9) (Amirreza Mahbod *et al*. 2019), use three pre-trained models of CNN in parallel which is AlexNet, VGG16and ResNet-18. These models are used to extract feature from skin lesion and then used to train SVM for classification. In total, the dataset used in this paper are 2037 colour dermoscopic skin images which include 411 malignant melanoma (MM), 254 seborrheic keratosis (SK) and 1372 benign nevi (BN). Normalization, resizing, rotation with difference angles and flipping are adjust as pre-processing and augmentation step before classification. The average accuracy is 90.69%.
10) (Anupama M A *et al*. 2019), present a model for classify 285 histology images of breast cancer using capsule network into four categories, benign, in situ, invasive and normal. Patch extraction and stain normalization are applied on the images as pre-processing step before the data fed into the model. The convolution layer used to extract the feature, after that capsule layer consists 51 capsules. Fully connected layer is the last layer of this architecture. The accuracy of this model is 92.14%.

Table 1 A compression for the models discussed in this paper

Models	Type of medical image	Optimizer methods	Epoch	Time	Tool	Accuracy %
(Hossam H.Sultan *et al*. 2019)	MRI of brain tumor	SGD	100	Study I: 289 minutes Study II: 2.5 minutes	Matlab & Python	Study I: 96.13 Study II: 98.7
(Amin Kabir *et al*. 2019)	MRI of brain tumor	Adam	100	Short time	-	Study I: 90.0 Study II:94.2
(Ali M. Hasan *et al*. 2019)	MRI of brain tumor	-	100	-	Matlab	99.30
(S. Sasikala *et al*. 2018)	MRI of brain tumor	-	100	-	Matlab	96
(Gabo Liang *et al*. 2018)	BCCD of blood cells	Adam	70	14 hours	Tensorflow	90.79
(Muhammad Sajjad *et al*, 2019)	MRI of brain tumor	-	30	-	Caffe	Study I: 90.67 Study II: 96.12
(Ashnil Kumar *et al*. 2016)	30 modalities of medical images	-	50	AlexNet: 14.722 s GoogLeNet: 39.394 s	Matlab	Top 1: 82.48 Top 5: 96.59
(Varun *et al*. 2019)	CT scan of lungs	-	200	-	Matlab	The average precision is 95.26 for both cases
(Amirreza Mahbod *et al*. 2019)	Skin lesion	-	-	-	-	90.69
(Anupama M A *et al*. 2019)	Histology images of breast cancer	-	40	-	Tensor flow	92.14

Conclusions

In this paper, we have discussed a summary of the basic concept of convolutional neural network and we have also surveyed some models that have appeared in recent years that address most of the problems in classifying various types of medical images using CNN as classifier or as feature extraction. However, we believe deep learning will likely remain an active area for research in the coming years in various applications, such as image recognition and classification, disease diagnosis, automatic translation and everything has an intelligence applications. So, many problems could be solved using deep learning better than the traditional methods.

References

LeCun, Y., *et al*. (1990)., "Handwritten digit recognition with a back-propagation network," In Proc. Advances in Neural Information Processing Systems, 396–404.

Krizhevsky, Alex (2013), "ImageNet Classification with Deep Convolutional Neural Networks,".

M. D. Zeiler and R. Fergus (2014), "Visualizing and understanding convolutional networks," in Computer Vision – ECCV 2014, D. Fleet, T. Pajdla, B. Schieleand T. Tuytelaars, Eds. Cham: Springer International Publishing, , pp. 818–833.

Simonyan, K., Zisserman, (2014), A.: Very deep convolutional networks for large-scale image recognition. CoRR, abs/1409.1556.

Farhana Sultana,Abu Sufian,Paramartha Dutta, (2018), 'Advancements in Image Classification using Convolutional Neural Network', 2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICNz).

Phil Kim. (2017) "MATLAB Deep Learning", Springer Science and Business Media LLC.

Yann LeCun, Yoshua Bengio, Geoffrey Hinton, 2015, 'Deep learning', Nuture, vol 5 2 1.

Ian Goodfellow, Yoshua Bengio, Aaron Courville, 'Deep Learning', Google books,
https://books.google.iq/books?id=Np9SDQAAQBAJ&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false

Reza Bosagh Zadeh, Bharath Ramsundar, TensorFlow for Deep Learning, https://www.oreilly.com/library/view/tensorflow-for-deep/9781491980446/ch04.html.

Nair, V., Hinton, G.E. (2010): Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th International Conference on International Conference on Machine Learning, ICML'10, pp. 807–814. USA. Omnipres

R. Ramani,N.Suthanthira Vanitha,S. Valarmathy, 2013, 'The Pre-Processing Techniques for Breast Cancer Detection in Mammography Images', International Journal of Image, Graphics and Signal Processing, vol. 5, no. 5, pp. 47-54

Bottou, L(2010): Large-scale machine learning with stochastic gradient descent. In: Lechevallier, Y.,Saporta, G. (eds.) Proceedings of COMPSTAT'2010, pp. 177–186. Heidelberg, Physica-Verlag HD.

Ruder, S. (2016): An overview of gradient descent optimization algorithms. CoRR, abs/1609.04747.

Hossam H. Sultan,Nancy M. Salem,Walid Al-Atabany, 2019, 'Multi-Classification of Brain Tumor Images Using Deep Neural Network', IEEE Access, vol. 7, pp. 69215-69225.

Amin Kabir Anaraki,Moosa Ayati,Foad Kazemi, 2019, 'Magnetic resonance imaging-based brain tumor grades classification and grading via convolutional neural networks and genetic algorithms', Biocybernetics and Biomedica Engineering, vol. 39, no. 1, pp. 63-74.

Ali M. Hasan,Hamid A. Jalab,Farid Meziane,Hasan Kahtan,Ahmad Salah Al-Ahmad, 2019, 'Combining Deep and Handcrafted Image Features for MRI Brain Scan Classification', IEEE Access, vol. 7, pp. 79959-79967.

S. Sasikala, M. Bharathi, B. R. Sowmiya, 2018, 'Lung Cancer Detection and Classification Using Deep CNN', International Journal of Innovative Technology and Exploring Engineering (IJITEE).

Gaobo Liang,Huichao Hong,Weifang Xie,Lixin Zheng, 2018, 'Combining Convolutional Neural Network With Recursive Neural Network for Blood Cell Image Classification', IEEE Access, vol. 6, pp. 36188-36197.

Muhammad Sajjad,Salman Khan,Khan Muhammad,Wanqing Wu,Amin Ullah,Sung Wook Baik, 2019, 'Multi-grade brain tumor classification using deep CNN with extensive data augmentation', Journal of Computational Science, vol. 30, pp. 174-182.

Ashnil kumar, David Lyndon, 2016, 'An Ensemble of Fine-Tuned Convolutional Neural Networks for Medical Image Classification', IEEE journal of biomedical and health informatics

Varun Srivastava,Ravindra Kr. Purwar, 2019, 'Classification of CT Scan Images of Lungs Using Deep Convolutional Neural Network with External Shape-Based Features', Journal of Digital Imaging.

Amirreza Mahbod,Gerald Schaefer,Chunliang Wang,Rupert Ecker,Isabella Ellinge, 2019, 'Skin Lesion Classification Using Hybrid Deep Neural Networks', ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

M.A. Anupama,V. Sowmya,K.P. Soman, 2019, 'Breast Cancer Classification using Capsule Network with Preprocessed Histology Images', 2019 International Conference on Communication and Signal Processing (ICCSP).